

## **Kahn et al. Data Quality Framework**

Adapted from: Kahn, Michael G.; Callahan, Tiffany J.; and Barnard, Juliana; et al. (2016) "**A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data,**" *eGEMs (Generating Evidence & Methods to improve patient outcomes): Vol. 4: Iss. 1, Article 18*. DOI:<http://dx.doi.org/10.13063/2327-9214.1244>.

Note that Table 1 in the Kahn et al. publication includes examples of the various elements of the framework.

- **Category 1: Conformance:** Data quality features that describe the compliance of the representation of data against internal or external formatting, relational, or computational definitions.
  - **Value Conformance:** Are data elements in agreement with a prespecified, constraint-driven data architecture?
  - **Relational Conformance:** Are data elements in agreement with additional structural constraints imposed by the physical database structures that store data values?
  - **Computational Conformance:** Do computations used to create derived values from existing variables yield the intended results either within a data set (Verification) or between data sets (Validation), when programs are based on identical specifications?
- **Category 2: Completeness:** Data quality features that describe the frequencies of data attributes present in a data set without reference to data values.
- **Category 3: Plausibility:** focuses on data quality features that describe the believability or truthfulness of data values. For this category, plausibility is determined by a variable's value.
  - **Uniqueness Plausibility:** Do objects (entities, observations, facts) appear multiple times in settings where they should not be duplicated or cannot be distinguished within a database (Verification) or when compared with an external reference (Validation)?
  - **Atemporal Plausibility:** Do observed data values, distributions, or densities agree with local or "common" knowledge (Verification) or from comparisons with external sources that are deemed to be trusted or relative gold standards (Validation).
  - **Temporal Plausibility:** Do time varying variables change values as expected based on known temporal properties or across one or more external comparators or gold standards?